

基于局部和全局特征提取优化的人脸图像修复^①



柏武斌^{1,2}, 张 乾^{1,2}

¹(贵州民族大学 数据科学与信息工程学院, 贵阳 550025)

²(贵州民族大学 贵州省模式识别与智能系统重点实验室, 贵阳 550025)

通信作者: 张 乾, E-mail: gzmuzq@gzmu.edu.cn

摘 要: 针对现有算法无法有效整合局部细节和全局结构的问题, 提出一种融合注意力机制优化局部和全局特征的三阶段人脸图像修复算法. 第 1 阶段引入位置注意力 (position attention module, PAM) 和聚焦线性注意力 (focused linear attention, FLA) 机制以增强图像局部纹理细节和全局上下文特征的提取. 在第 2 阶段优化中, 为了提升局部细节的修复效果, 引入卷积注意力模块 (convolutional block attention module, CBAM), 并结合跳跃连接机制. 该设计通过通道和空间维度的差异化权重分配强化特征关注, 同时利用下采样过程中的细节保留策略, 有效实现局部区域的精细化重建. 最后, 引入第 3 阶段整合特征, 使修复图像更具有鲁棒性. 实验结果表明, 该方法在 CelebA-HQ 数据集上 PSNR 和 SSIM 平均提高了 0.1214 dB 和 0.0022, LPIPS 平均下降了 0.00065, 显著提高了修复图像质量和视觉效果.

关键词: 位置注意力模块; 聚焦线性注意力机制; CBAM; 感受野; 通道和空间

引用格式: 柏武斌, 张乾. 基于局部和全局特征提取优化的人脸图像修复. 计算机系统应用, 2025, 34(8): 139-148. <http://www.c-s-a.org.cn/1003-3254/9945.html>

Face Image Restoration Based on Local and Global Feature Extraction Optimization

BAI Wu-Er^{1,2}, ZHANG Qian^{1,2}

¹(School of Data Science and Information Engineering, Guizhou Minzu University, Guiyang 550025, China)

²(Key Laboratory of Pattern Recognition and Intelligent Systems of Guizhou, Guizhou Minzu University, Guiyang 550025, China)

Abstract: To address the problem that existing algorithms cannot effectively integrate local details and global structures, this study proposes a three-stage face image restoration algorithm that incorporates attention mechanisms to optimize local and global features. In the first stage, the position attention module (PAM) and focused linear attention (FLA) are introduced to enhance the extraction of local texture details and global contextual features of the image. In the second stage of optimization, the convolutional block attention module (CBAM) is incorporated with skip connections. The design strengthens feature focus through the differentiated weight assignment in both channel and spatial dimensions, while effectively achieving refined reconstruction of local regions by utilizing the detail-preservation strategy during the downsampling process. Finally, the third stage of integrating features is introduced to make the restored image more robust. The experimental results show that the proposed method achieves average improvements of 0.1214 dB in PSNR and 0.0022 in SSIM, along with an average reduction of 0.00065 in LPIPS on the CelebA-HQ dataset, significantly enhancing both the restoration quality and visual perception of images.

Key words: position attention module (PAM); focused linear attention (FLA) mechanism; convolutional block attention module (CBAM); receptive field; passage and space

① 基金项目: 贵州省高等学校大数据分析与应用重点实验室(黔教技[2023]012号); 贵州民族大学校级科研项目(GZMUZK [2021] YB23, GZMUZK [2023] QN10)

收稿时间: 2024-12-25; 修改时间: 2025-02-12; 采用时间: 2025-03-06; csa 在线出版时间: 2025-06-20

CNKI 网络首发时间: 2025-06-23

图像修复 (image inpainting)^[1]是指修复或恢复图像中的损坏部分或缺失区域。它的目标是通过分析图像的上下文信息,算法可以生成与周围区域一致的填补内容,从而恢复图像的完整性,使得修复后的图像看起来尽可能自然和连贯,保持视觉的一致性和真实感。图像修复广泛应用于照片修复、艺术品修复以及视频处理等领域,旨在消除瑕疵或去除不需要的物体。现有的图像修复方法可分为基于传统方法和深度学习方法。

传统修复方法主要是基于扩散^[1,2]和基于补丁^[3-5]的方案,其依赖于视觉低级假设和结构来搜索相似的补丁以进行重建,而不能处理不规则或复杂的掩模。为了解决此问题,大量注意力转移至深度学习的方法,如卷积神经网络 (convolutional neural network, CNN)^[6]和对抗生成网络 (generative adversarial network, GAN)^[7],按照网络模型设计可将其修复方法大致分为单阶段修复网络、双阶段修复网络、渐进式修复网络和基于注意力的修复网络。

(1) 单阶段修复网络:单阶段修复网络较为简单,它们可用于快速修复图像中的小瑕疵或噪声。如 Pathak 等人^[8]设计了一种编码-解码器架构,该架构通过像素重建损失和对抗性损失的组合进行训练。为了提高图像补全的一致性, Iizuka 等人^[9]引入了全局和局部上下文判别器来训练全卷积补全网络。为了更好地训练网络, Hui 等人^[10]设计了特征中心对齐约束,并设计了一个自引导回归损失来增强语义细节,但在修复中不能区分有效像素和无效像素,导致计算冗余,而 Wang 等人^[11]提出了有效性可迁移卷积和区域复合归一化模块,以在修复过程中更好地利用有效像素。

单阶段图像修复方法细节推理速度快,适合简单图像修复,但修复质量较低,处理复杂区域和细节时效果较差。

(2) 双阶段修复网络:为了改善一阶段的修复缺陷, Nazeri 等人^[12]提出了一种边缘引导的两阶段图像修复方法,通过恢复了缺失区域的边缘图,然后将这个边缘图与不完整的图像结合起来作为第2阶段的输入来执行修复任务。由于文献^[12]边缘图像的结构指导有限, Ren 等人^[13]采用边缘保持平滑处理的图像作为结构的表示,然后将修复任务建模为结构重建和纹理生成的组合。Wu 等人^[14]引入 LBP 特征来捕捉图像的局部纹理信息,同时利用空间注意力机制来聚焦图像中重要的区域,从而提升局部和全局的修复。

双阶段修复方法在全局结构和局部细节修复上都有较好表现,但修复精度和计算开销较高,对于大范围损坏和复杂场景的恢复较差。

(3) 渐进式修复网络:为了渐进地重建视觉结构, Guo 等人^[15]设计残差架构促进特征整合和纹理预测,采用 n 块、单扩张策略,并应用步长损失函数来提高中间修复的性能。Li 等人^[16]在特征空间中遵循渐进式框架,设计了具有一致注意力的递归特征推理网络,递归机制帮助网络在多次推理过程中逐步恢复更细致的纹理和结构,但处理复杂纹理或结构的地方时,边界的修复效果不如预期。Zeng 等人^[17]提出了一种基于置信度的迭代修复方法,该网络类似于文献^[18],第2个生成器的输出是修复图像沿着置信度图,并且该图用于指导完成过程的下一次迭代。Dong 等人^[19]提出掩蔽位置编码与 Transformer 结构恢复器,用增量式训练和多组件协作,提升具有复杂结构图像的修复效果。Quan 等人^[20]提出了具有不同大小感受野的局部和全局精细化网络,有效改善大面积遮挡图像,但对局部边缘修复不够自然。

渐进式图像修复网络通过逐步细化修复,能更好地恢复复杂全局结构与局部细节,特别适合大范围损坏,但计算复杂度高,推理速度较慢。整体而言,渐进式方法在全局特征修复上优势更明显。

(4) 基于注意力的图像修复方法

Yu 等人^[18]提出了一种基于上下文注意力机制的生成图像修复方法,通过上下文信息动态地聚焦于缺失区域的周围区域,有助于填补细节区域,修复自然纹理,减少伪影和边界不一致问题,但是在复杂背景或结构化图像中,修复效果不如局部细节精细。Oktay 等人^[21]在 U-Net 结构中结合了注意力机制,使模型更加关注输入图像中的重要区域,避免无关区域的干扰,对纹理复杂或边缘区域有着较好的修复效果,但对于小尺寸图像或者非常细微的局部修复问题不够敏感,导致修复效果不够细腻。Fu 等人^[22]提出了双重注意力网络 (DANet), 将其空间注意力和通道注意力机制结合,有效提高局部纹理细节恢复方面,但在处理复杂背景或大面积缺失的区域时,修复不够自然。Qin 等人^[23]提出了一种多尺度注意力模块,通过合并不同大小 patch 的注意力模块产生的特征图,来捕捉多个尺度的信息,提升局部区域与周围原始区域的纹理过渡自然,但难以从全局角度对图像的结构、色彩等进行统一协调。

综上注意力机制能够在修复过程中重点关注图像中的局部细节区域,减少边缘模糊和过度平滑等现象,但是过于依赖局部上下文,导致修复结果在全局层面缺乏一致性,尤其是在大范围缺失或背景复杂的情况下。

本文尝试融合渐进式修复网络与注意力机制,以实现图像局部与全局特征的深度优化:(1)渐进式修复网络凭借其分阶段、逐步细化的修复策略,展现出卓越的应对复杂场景及大面积缺失区域图像修复的能力,它能够依据图像周边环境及已修复部分的信息,有条不紊地重建缺失内容,确保修复后的图像在整体布局、结构逻辑上的合理性;(2)巧妙地在渐进式修复的各个关键阶段嵌入多样化的注意力机制,精准聚焦图像的局部微观特征,例如人脸图像中的睫毛细节、人脸中的纹理细节等,借助注意力的引导,让模型在修复进程中将这些细微之处还原得更加清晰、精准,全方位提升局部修复的精细度与逼真感,从而实现从局部细节到全局布局的高质量图像修复效果。

1 算法框架

本文算法框架如图1所示,该框架是由典型的U-Net和卷积神经网络结构组成,结合了不同感受野大小来生成缺失区域。它由3个网络组成,分别为第1-3阶段网络,3个网络逐次连接进行训练。

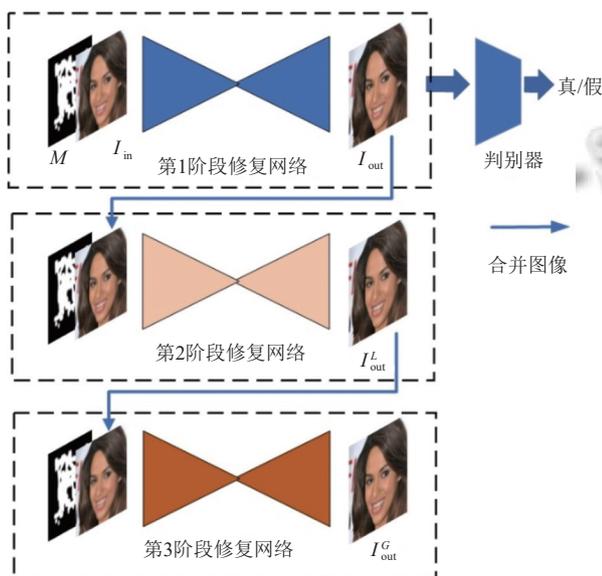


图1 3阶段网络框架图

第1阶段网络是典型的U-Net结构,使用深层网络扩大感受野进行修复训练,以捕获长距离依赖关系

的像素并生成全局合理的整体形状和语义结构。例如头发、整体形状轮廓等,但是在修复局部的纹理细节,边缘等效果比较差,如眼睛、嘴唇等。为此,引入第2阶段网络,通过浅层的卷积神经网络来缩小感受野进行局部修复,使局部的纹理细节得到更好的修复。引入第3阶段进行特征全局精化,使修复结果更具有鲁棒性。

1.1 第1阶段修复网络

第1阶段修复网络结构(图2)由判别器和生成器组成。生成器主要由8个下采样,PAM-FLA注意力模块,和8个上采样组成,用于生成逼真的图像。判别器主要作用是辨别生成器生成图像的真伪性,两者相互博弈,使得生成器生成的图像能够“骗过”判别器,从而生成高质量的图片。

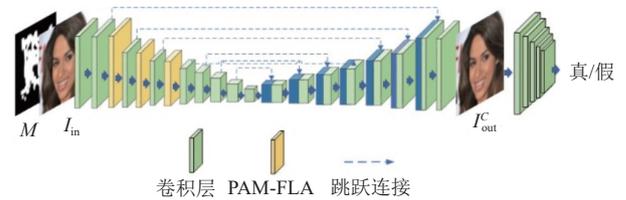


图2 第1阶段修复网络

将掩码图像 M (其中0表示有效像素,1表示缺失像素)和原始图像 I_{in} 作为输入数据,经过第1修复网络得到初步修复结果 I_{out}^C 。其中 I_{in} 和掩码 M 融合生成 I_{mer}^C 作为第2阶段网络的输入数据,具体定义如下:

$$I_{out}^C = d^8(c^5(pf(c(pf(c(pf(c(M, I_{in})))))))) \quad (1)$$

$$I_{mer}^C = I_{in} \odot (1 - M) + I_{out}^C \odot M \quad (2)$$

其中, c 为卷积层, c^5 表示5个卷积层, pf 为PAM-FLA模块, d^8 为8个反卷积, \odot 表示逐元素乘积运算。

1.1.1 PAM-FLA 模块

初始阶段的修复网络通常依赖深层架构来着力捕捉图像的全局上下文信息,然而这一做法会导致进入第2阶段修复时,可获取的正确信息匹配大幅减少。鉴于此,在此阶段下采样环节引入了PAM-FLA注意力模块。该模块功能卓越,一方面能够精准捕获局部精细的纹理细节,另一方面可以有效把握图像全局长远距离的内在关系,极大程度地降低了下采样过程中的信息丢失风险,进而让第1阶段的修复工作得以同时、均衡地聚合局部与全局两类关键特征。

PAM-FLA模块是由位置注意力机制模块(position attention module, PAM)^[24]和聚焦线性注意力机制

(focused linear attention, FLA)^[25]组成,能够很好地修复破损区域的局部纹理细节和全局整体结构,对大面积

破损结构有着良好的修复效果,两者结合能够提高修复效果的精确度和自然度,其结构如图3所示.

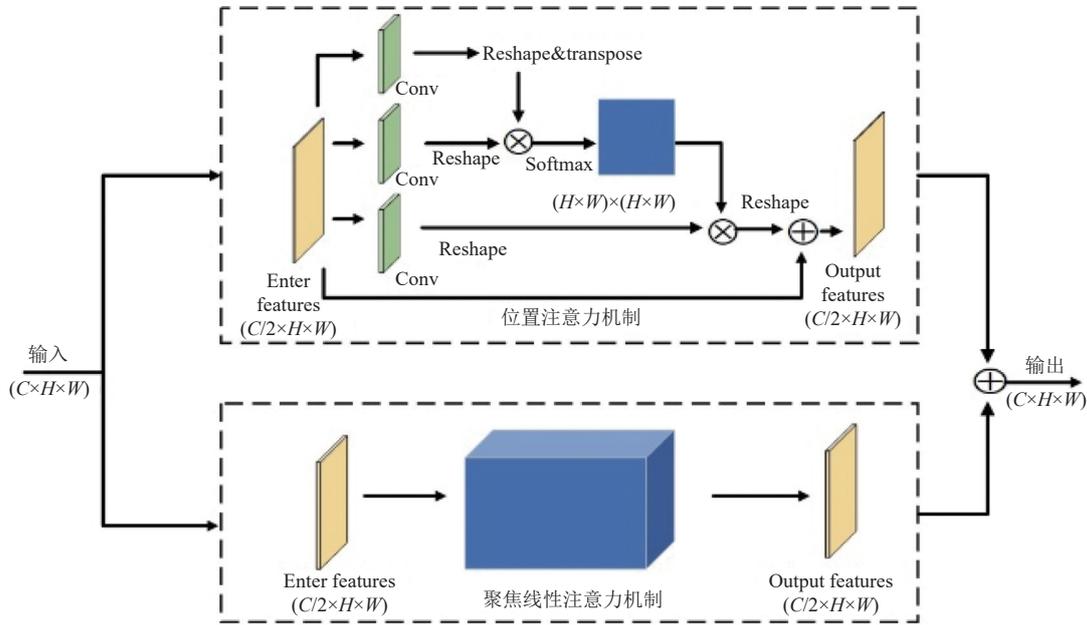


图3 PAM-FLA 结构图

首先, PAM 能够显式地关注输入图像或特征图 X 的不同空间位置, 以增强模型对图像中局部细节的感知能力, 从而提高局部修复的精度. 此外, 能够根据不同任务和数据集自动调整注意力权重, 对于复杂的场景和变化的输入数据也能取得良好效果. 定义如下:

$$X'_p = T(R(c(X))) \quad (3)$$

$$X''_p = R(c(X)) \otimes X'_p \quad (4)$$

$$X'''_p = R(c(X)) \otimes \delta(X''_p) \quad (5)$$

$$X_p = X'''_p \oplus X \quad (6)$$

其中, R 表示重塑形状函数, T 表示转置函数, δ 表示 Softmax 激活函数, X_p 表示经过位置注意力机制处理后输出的特征.

FLA 在处理大掩码比例结构图像或者需要修复复杂结构的场景中. 它能够帮助模型集中注意力于图像中的关键区域, 以便更准确地恢复缺失的信息. 此外, 可以通过对不同位置的加权来确保修复的部分与周围环境保持一致. 这有助于生成更加自然和逼真的修复结果, 避免明显的修复痕迹. FLA 可以帮助模型在保持局部细节的同时, 有效地处理远距离的上下文信息,

提高修复的精度和完整性, 定义如下:

$$X_f = FLA(X) \quad (7)$$

其中, X_f 表示经过聚焦线性注意力机制 FLA 处理后输出的特征, 结合两特征进行特征增强, 定义如下:

$$X_{out} = X_f \oplus X_p \quad (8)$$

X_{out} 特征保持良好的全局一致性, 同时保留和恢复局部的细节信息, 使修复结果更加自然和逼真. 在应对复杂的损坏类型结构和场景 (如遮挡、大面积损失等) 时, 可以灵活应对, 提升修复效果的鲁棒性和通用性.

1.2 第2阶段修复网络

第2阶段网络使用卷积神经网络 (图4) 进行局部修复. 该网络由两个下采样, 残差块, CBAM 模块^[26], 两个跳跃连接和两个上采样组成.

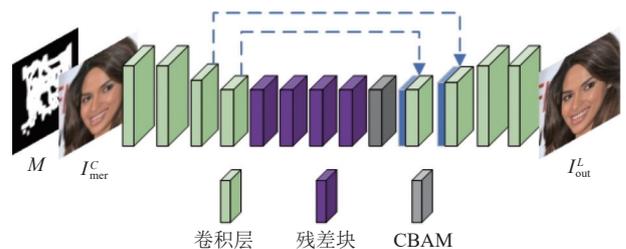


图4 第2阶段修复网络

掩码 M 和 I_{mer}^C 作为输入数据, 生成局部修复结果 I_{out}^L 与原始图像 I_{in} 和掩码 M 融合生成 I_{mer}^L , 其作为第3阶段的输入数据, 具体定义如下:

$$I_{out}^L = d^3 \left(cbam \left(r^4 \left(c^3 \left(M, I_{mer}^C \right) \right) \right) \right) \quad (9)$$

$$I_{mer}^L = I_{in} \odot (1 - M) + I_{out}^L \odot M \quad (10)$$

其中, $cbam$ 为 CBAM 注意力机制, r^4 为 4 个残差块。

1.2.1 CBAM 注意力机制

为了增强局部特征表示, 并且提取特征过程注重重要区域和通道, 在第2网络中引入了跳跃连接和 CBAM 注意力机制。CBAM 能够动态地调整对重要区域和通道的关注程度, 从而增强特征表达能力, 提高模型的修复性能。

CBAM 模块主要由两个子模块组成: 通道注意力模块和空间注意力模块。这两个子模块以串行的方式组合在一起, 序列化的在通道维度和空间维度上产生注意力特征图信息。利用通道间的特征关系生成通道注意力图, 进而连接到空间注意力得到富含信息量的特征图, 具体结构如图5所示。相关计算表示如下:

$$F' = M_c(F) \otimes F \quad (11)$$

$$F'' = M_s(F') \otimes F' \quad (12)$$

其中, F 表示输入特征, F' 表示通过通道注意力获得的特征, F'' 表示经过空间注意力后得到的特征, $M_c(\cdot)$ 表示经过池化后生成的一维通道注意力, $M_s(\cdot)$ 表示经过池化后生成的二维空间注意力。

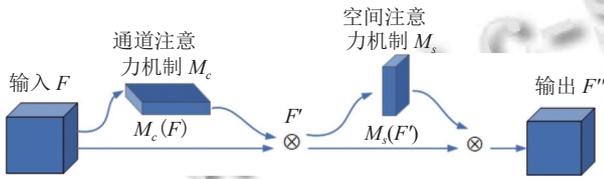


图5 CBAM 注意力机制结构图

1.3 第3阶段修复网络

为进一步优化图像修复效果, 提升全局结构的修复质量, 增强修复结果的鲁棒性, 引入了第3阶段全局精细化网络^[20]。该网络主要由8个下采样层、8个上采样层以组成(图6), 从全局视角对图像进行深度精化, 让修复后的图像在整体结构完整性与稳定性上达到更优水准。

将掩码 M 和 I_{mer}^L 作为输入数据, I_{out}^G 为该阶段修复

结果。其中 I_{out}^G 、 M 和 I_{in} 进行融合生成该阶段最终的修复结果 I_{out} , 相关定义如下:

$$I_{out}^G = d^8 \left(c^8 \left(M, I_{out}^L \right) \right) \quad (13)$$

$$I_{out} = I_{in} \odot (1 - M) + I_{out}^G \odot M \quad (14)$$

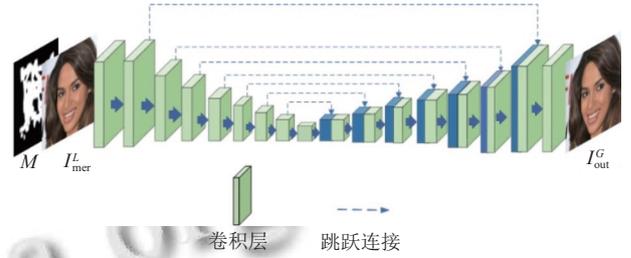


图6 第3阶段修复网络

1.4 损失函数

本实验的训练损失函数遵循文献[20]的损失函数。

(1) 第1阶段损失函数

本阶段训练目标包括像素重建损失和对抗损失, 在本阶段引入 L1 损失作为像素级重建损失, 公式如下:

$$L_{valid}^C = \frac{1}{\sum(1 - M)} \left\| (I_{out}^C - I_{gt}) \odot (1 - M) \right\|_1 \quad (15)$$

$$L_{hole}^C = \frac{1}{\sum(1 - M)} \left\| (I_{out}^C - I_{gt}) \odot M \right\|_1 \quad (16)$$

其中, I_{gt} 表示真实的图像数据, $\sum(M)$ 表示 M 中非零元素的数量。L1 损失公式化如下:

$$L_r^C = L_{valid}^C + \lambda_h \cdot L_{hole}^C \quad (17)$$

其中, h 为平衡因子。对抗损失函数有生成器损失和判别器损失, 两者都使用最小二乘损失, 公式如下:

$$L_G^C = E_{I_{mer} \sim p_{I_{mer}}} \left[(D(I_{mer}^C) - 1)^2 \right] \quad (18)$$

$$L_D = \frac{1}{2} E_{I \sim p_{data}(I)} \left[(D(I_{gt}) - 1)^2 \right] + \frac{1}{2} E_{I_{mer} \sim p_{I_{mer}}} \left[(D(I_{mer}^C))^2 \right] \quad (19)$$

其中, L_G^C 为生成器损失, L_D 为判别器损失, 该阶段的损失函数记为:

$$L_C = L_{valid}^C + \lambda_h \cdot L_{hole}^C + \lambda_g \cdot L_G^C \quad (20)$$

本次实验设置 $\lambda_h = 6$, $\lambda_g = 0.1$ 。

(2) 第2阶段损失函数

本阶段损失函数由重建损失、总变分损失、感知损失和风格损失组成。重建损失依然采用 L1 损失, 公

式如下:

$$L_{\text{valid}}^L = \frac{1}{\text{sum}(1-M)} \|(I_{\text{out}}^L - I_{\text{gt}}) \odot (1-M)\|_1 \quad (21)$$

$$L_{\text{hole}}^L = \frac{1}{\text{sum}(1-M)} \|(I_{\text{out}}^L - I_{\text{gt}}) \odot M\|_1 \quad (22)$$

$$L_r^L = L_{\text{valid}}^L + \lambda_h \cdot L_{\text{hole}}^L \quad (23)$$

总变分损失函数表示为:

$$L_{\text{tv}}^L = \|I_{\text{mer}}^L(i, j+1) - I_{\text{mer}}^L(i, j)\|_1 + \|I_{\text{mer}}^L(i+1, j) - I_{\text{mer}}^L(i, j)\|_1 \quad (24)$$

其中, I_{mer}^L 表示第 2 阶段输出的图像数据 I_{out}^L 与 M 的融合图像, 其计算方法和式 (2) 类似. 感知损失函数公式为:

$$L_{\text{per}}^L = \sum_i \|F_i(I_{\text{out}}^L) - F_i(I_{\text{gt}})\|_1 + \|F_i(I_{\text{mer}}^L) - F_i(I_{\text{gt}})\|_1 \quad (25)$$

其中, F_i 表示预训练 VGG-16 网络中第 i 层的特征图. 风格损失函数公式为:

$$L_{\text{sty}}^L = \sum_i \|G_i(I_{\text{out}}^L) - G_i(I_{\text{gt}})\|_1 + \|G_i(I_{\text{mer}}^L) - G_i(I_{\text{gt}})\|_1 \quad (26)$$

其中, G_i 表示格拉姆矩阵. 所以本阶段的损失函数 L_L 为:

$$L_L = L_{\text{valid}}^L + \lambda_h \cdot L_{\text{hole}}^L + \lambda_{\text{tv}} \cdot L_{\text{tv}}^L + \lambda_{\text{per}} \cdot L_{\text{per}}^L + \lambda_{\text{sty}} \cdot L_{\text{sty}}^L \quad (27)$$

根据文献[14], 设置相关参数: $\lambda_h = 6$, $\lambda_{\text{tv}} = 0.1$, $\lambda_{\text{per}} = 0.05$ 和 $\lambda_{\text{sty}} = 120$.

(3) 第 3 阶段损失函数

本阶段的损失函数和第 2 阶段损失函数相似. 重建损失函数为:

$$L_{\text{valid}}^G = \frac{1}{\text{sum}(1-M)} \|(I_{\text{out}}^G - I_{\text{gt}}) \odot (1-M)\|_1 \quad (28)$$

$$L_{\text{hole}}^G = \frac{1}{\text{sum}(1-M)} \|(I_{\text{out}}^G - I_{\text{gt}}) \odot M\|_1 \quad (29)$$

$$L_r^G = L_{\text{valid}}^G + \lambda_h \cdot L_{\text{hole}}^G \quad (30)$$

总变分损失函数表示为:

$$L_{\text{tv}}^G = \|I_{\text{mer}}^G(i, j+1) - I_{\text{mer}}^G(i, j)\|_1 + \|I_{\text{mer}}^G(i+1, j) - I_{\text{mer}}^G(i, j)\|_1 \quad (31)$$

其中, I_{mer}^G 表示第 2 阶段输出的数据 I_{out}^G 与 M 的融合图像. 感知损失函数公式为:

$$L_{\text{per}}^G = \sum_i \|F_i(I_{\text{out}}^G) - F_i(I_{\text{gt}})\|_1 + \|F_i(I_{\text{mer}}^G) - F_i(I_{\text{gt}})\|_1 \quad (32)$$

其中, F_i 表示预训练 VGG-16 网络中第 i 层的特征图. 风格损失函数表示为:

$$L_{\text{sty}}^G = \sum_i \|G_i(I_{\text{out}}^G) - G_i(I_{\text{gt}})\|_1 + \|G_i(I_{\text{mer}}^G) - G_i(I_{\text{gt}})\|_1 \quad (33)$$

其中, G_i 表示格拉姆矩阵, 则本阶段的损失函数 L_G 为:

$$L_G = L_{\text{valid}}^G + \lambda_h \cdot L_{\text{hole}}^G + \lambda_{\text{tv}} \cdot L_{\text{tv}}^G + \lambda_{\text{per}} \cdot L_{\text{per}}^G + \lambda_{\text{sty}} \cdot L_{\text{sty}}^G \quad (34)$$

相关参数设置同上.

总的损失函数由 3 个子网络的损失函数和判别器损失函数构成, 即:

$$L_{\text{all}} = L_C + L_L + L_G + L_D \quad (35)$$

2 实验

2.1 实验数据与实验环境

本文算法模型在高清人脸 CelebA-HQ^[27]数据集和不规则掩码数据集^[28]下进行实验. 其中 CelebA-HQ 包含了 30 000 张尺寸为 256×256 的高清人脸图像, 按照 1:14 的比例, 将数据集前 28 000 张数据用于训练部分, 后 2 000 张数据用于测试部分. 不规则掩码包含 12 000 张尺寸为 256×256 的数据, 按照掩码受损率将掩码数据分为 6 类不同比例, 分别为: 0–10%, 10%–20%, 20%–30%, 30%–40%, 40%–50%, 50%–60%. 图 7 所示为 6 种不同掩码比率遮挡效果图.

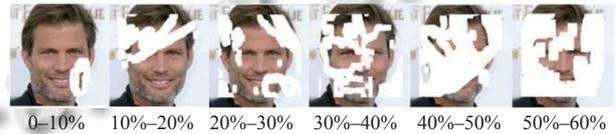


图 7 6 种不同掩码比率遮挡效果图

实验基于 PyTorch 1.13, Python 3.8 和 CUDA 11.8 环境下, 在 GPU 为 NVIDIA GeForce RTX 3090 上进行训练和测试. 模型网络参数 Batch size 设置为 4, 利用 Adam 算法优化模型参数, 一共迭代了 200 轮训练, 前 100 轮学习率为 0.000 2, 后 100 轮线性下降至 0, 其中衰减指数 $\beta_1 = 0.5$ 和 $\beta_2 = 0.999$.

2.2 实验结果与分析

为了验证该网络模型的有效性, 将本文算法网络与 LGNet^[20]、CTSDG^[29]、MISF^[30]和 T-former^[31]算法在同一数据集上进行定量和定性实验分析.

2.2.1 定量实验

定量实验分析中选取图像修复领域的 3 个常用衡

量指标: 峰值信噪比 (peak signal-to-noise ratio, PSNR)、结构相似性 (structural similarity, SSIM) 和感知相似度 (learned perceptual image patch similarity, LPIPS). 其中 PSNR 反映了两幅图像在像素级别上的差异程度, 值越高, 表示两幅图像越相似, 修复效果越好. SSIM 是基于人类视觉系统的特点, 通过比较亮度、对比度和结构 3 个方面来评估图像的相似性, 取值范围从-1 到 1, 值越接近 1, 表示两幅图像越相似. 而 LPIPS 反映两幅图

的相似性, 值小说明修复质量越好.

为了更精准呈现不同掩码比例下各算法的修复量化指标表现, 进行如下分析. 由表 1 可知, 在相同破损率情况下, 与其余 4 种对比方法相比, 本文所提出的算法展现出卓越优势. 就 PSNR 和 SSIM 两项关键指标而言, 本文算法在各破损率条件下均达到最优的状态; LPIPS 指标方面, 除掩码比率 10%–20% 区间外, 其余区间同样达到最优, 彰显出算法的稳定性与优越性.

表 1 不同方法在 CelebA-HQ 下的定量实验

指标	Masks	0–10%	10%–20%	20%–30%	30%–40%	40%–50%	50%–60%
PSNR↑ (dB)	CTSDG	39.8969	33.7144	30.2016	27.6726	25.6665	22.6989
	MISF	38.9752	32.9635	29.6130	27.1759	25.1938	22.2591
	T-former	36.9100	33.6780	30.4517	27.9717	25.8645	22.1514
	LGNet	40.8510	34.8505	31.1845	28.5085	26.4295	23.2926
	本文算法	41.0743	35.0093	31.3147	28.6067	26.5103	23.3296
SSIM↑	CTSDG	0.9862	0.9604	0.9256	0.8856	0.8407	0.7686
	MISF	0.9839	0.9558	0.9198	0.8795	0.8339	0.7612
	T-former	0.9801	0.9585	0.9266	0.8898	0.8466	0.7669
	LGNet	0.9872	0.9654	0.9347	0.8981	0.8564	0.7847
	本文算法	0.9878	0.9666	0.9367	0.9007	0.8598	0.7883
LPIPS↓	CTSDG	0.0062	0.0180	0.0350	0.0559	0.0809	0.1281
	MISF	0.0064	0.0178	0.0334	0.0517	0.0742	0.1169
	T-former	0.0108	0.0108	0.0290	0.0455	0.0670	0.1211
	LGNet	0.0051	0.0143	0.0275	0.0438	0.0635	0.1033
	本文算法	0.0048	0.0137	0.0266	0.0429	0.0626	0.1030

注: ↑表示越高越好, ↓表示越低越好.

进一步与 LGNet 算法展开对比, 在掩码比率 0–30% 的小掩码比率区间内, 本文算法的 PSNR 指标平均提升幅度达到 0.17 dB, SSIM 指标平均提升 0.0013, LPIPS 指标平均降低 0.0006, 这意味着在局部细微破损修复场景中, 本文算法不仅能更精准地还原图像纹理细节, 还能在感知层面贴近原始图像, 减少修复痕迹. 当掩码比率扩大至 30%–60% 的大掩码区间时, 本文算法的 PSNR 平均提升 0.072 dB, SSIM 平均提升 0.0032, LPIPS 平均降低 0.0007, 充分表明即便面对大面积破损, 本文算法依旧能够在全局层面保障修复后图像的整体结构连贯性, 同时兼顾局部纹理的细腻还原, 成功实现对局部纹理细节特征提取与全局上下文特征提取的双重优化, 赋予算法网络生成高度还原、纹理细腻且结构完整图像的卓越性能.

2.2.2 可视化分析

为了更直观的呈现本文算法的优越性, 绘制不同算法修复效果对比图如图 8 所示, 左上角分图为 SSIM 评价指标对比图; 左下角分图为 LPIPS 评价指标对比图; 右侧分图表示 PSNR 对比图 (左侧两个分图纵坐标

表示对应指标值, 横坐标表示不同掩码比率; 右侧分图横坐标表示 PSNR 值, 纵坐标表示不同掩码比率). 图 8 中, 从 SSIM 指标相应的折线图走势来看, 本文算法的折线图均高于其他算法, 表明本文算法在结构层面保持高度的一致性. 从 LPIPS 指标分析图像的色彩与纹理呈现效果, 发现除了在特定的掩码区间 10%–20% 之外, 在其余各个掩码区间内, 本文算法所生成的图像在色彩饱和度、纹理清晰度等方面均表现优越, 修复结果更为逼真、细腻, 说明本文算法能够精准捕捉并还原原始图像中的丰富视觉信息. 在 PSNR 指标中, 在各个掩码区间本文算法条形图均高于其他算法条形图, 说明本文算法能够有效地保留和恢复图像中的细节信息, 大大程度减少修复过程中的信息损失.

2.2.3 定性实验

图 9 所示为 6 种不同掩码比率下不同算法修复结果对比图. 从第 1–6 行掩码比率逐渐变大 (0–60%), (a) 为破损图像, (b)–(f) 列为 5 种不同方法的修复结果, (h) 为原始图像. 在 20%–40% 下, 对比 5 种方法发现, 本文整体修复效果较好, 特别在局部纹理细节, 而其他

方法修复结果比较偏亮、模糊等,例如第4行.在40%–60%下,可以清晰看到本文对整体轮廓修复较为正确,对长距离关系的像素修复较为友好,而其他4种方法

整体修复较为模糊,并且出现亮度不一致等问题.综上所述本文算法生成的图像相较其他算法局部纹理细节和全局整体结构修复效果更清晰,修复视觉效果更好.

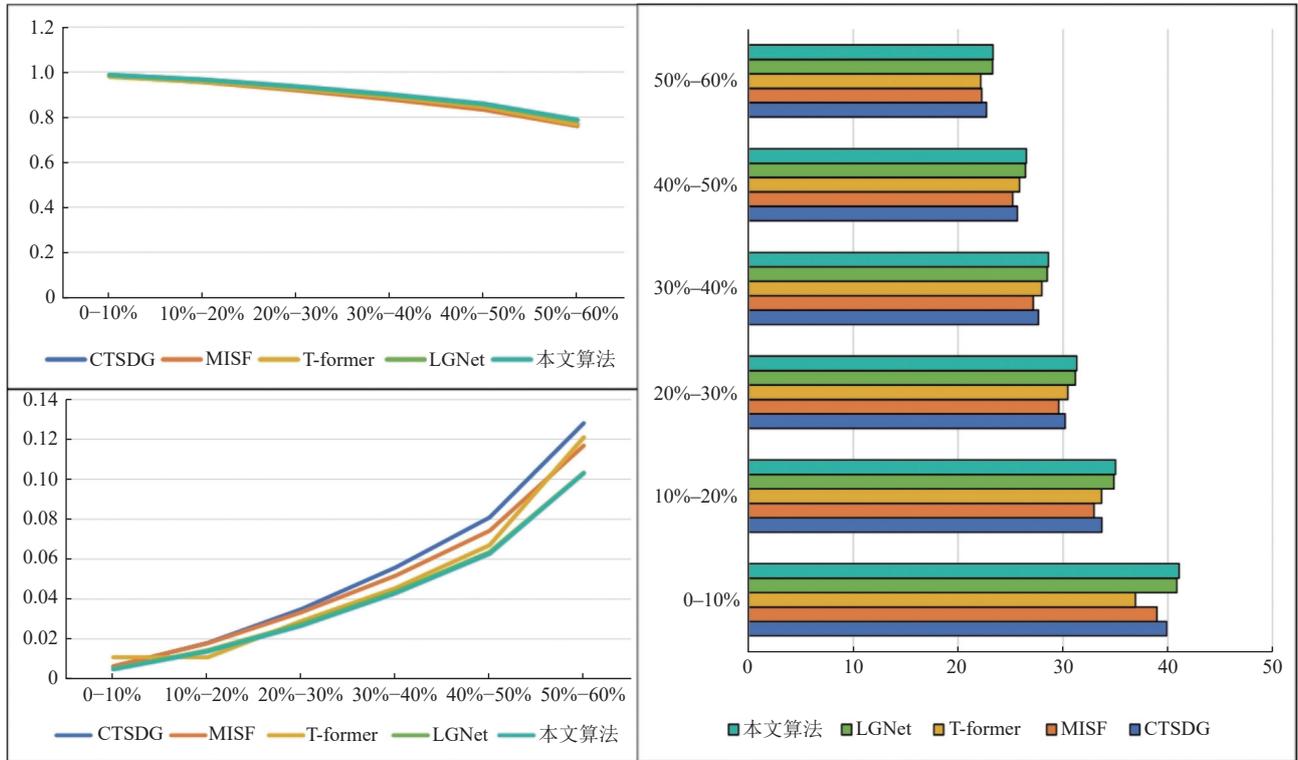


图8 不同算法对比效果图



图9 本文算法与其他方法对比效果图

2.3 消融实验

为了验证所提出模块的有效性,在同一数据集,相同配置环境下对各个模块进行消融实验.设置了4组包含不同模块的网络,如表2所示.

表2 模块消融实验客观指标

名称	PAM-FLA	CBAM	跳跃连接
网络1	×	×	×
网络2	√	×	×
网络3	√	√	×
网络4(本文算法)	√	√	√

注: ×表示未加该模块, √表示增加该模块

网络1表示没有添加任何模块;网络2表示在网络1的基础上添加 PAM-FLA 模块;网络3表示在网络2的基础上添加 CBAM 模块;网络4(本文算法)表示在网络3的基础上添加两个跳跃连接.

消融实验结果如表3所示,其中加粗的数据为最优网络.

由表3可知,在 PSNR 指标中,本文算法比网络1平均提高了 0.1214 dB.在掩码比率为 0–30%,随着模块的增加该指标提升较为明显,平均贡献率达到 70%,而在掩码率 30%–60%,该指标平均提高了 0.072 dB,平均贡献率达到 30%,说明在大掩码比率下,本模型修复效果较好.在 SSIM 指标中,该指标均有所提升,平

均提高了 0.25%, 而 LPIPS 指标值均有所下降, 平均下降了 1.51%。综上所述本文算法加入 PAM-FLA、CBAM

和跳跃连接均能够优化局部和全局的特征提取, 能够有效整合局部和全局特征。

表 3 不同模块消融实验对比

指标	Masks	0-10%	10%-20%	20%-30%	30%-40%	40%-50%	50%-60%
PSNR↑ (dB)	网络1	40.8510	34.8505	31.1845	28.5085	26.4295	23.2926
	网络2	40.9613	34.885	31.2175	28.5427	26.446	23.2723
	网络3	41.0421	34.9401	31.2469	28.5277	26.4456	23.2716
	本文算法	41.0743	35.0093	31.3147	28.6067	26.5103	23.3296
SSIM↑	网络1	0.9872	0.9654	0.9347	0.8981	0.8564	0.7847
	网络2	0.9877	0.9664	0.9363	0.9002	0.8590	0.7875
	网络3	0.9878	0.9665	0.9363	0.9000	0.8588	0.7871
	本文算法	0.9878	0.9666	0.9367	0.9007	0.8598	0.7883
LPIPS↓	网络1	0.0051	0.0143	0.0275	0.0438	0.0635	0.1033
	网络2	0.0049	0.0138	0.0268	0.0430	0.0629	0.1033
	网络3	0.0049	0.0139	0.0270	0.0435	0.0633	0.1039
	本文算法	0.0048	0.0137	0.0266	0.0429	0.0626	0.1030

注: ↑表示越高越好, ↓表示越低越好。

3 结论

本文引入了 3 阶段渐进式修复网络, 通过注意力机制优化第 1、2 阶段修复网络。首先第 1 阶段网络在下采样中并联式融合了 PAM 和 FLA 进行修复, 增强局部和全局信息的捕获能力, 减少局部信息丢失。在第 2 阶段引入了 CBAM 和跳跃连接进行局部优化, 通过自适应地强调重要特征并抑制不重要特征和生成细节, 从而提升模型的表现, 结合通道注意力和空间注意力, 使模型更好地聚焦于关键信息, 从而增强特征表达能力, 改善图像目标区域与周围区域边界不连贯的问题, 同时减少在下采样过程中信息丢失。为了使得图像在大掩码比率下修复更具有鲁棒性和符合视角审美效果, 引入第 3 阶段修复网络。实验结果证明本文算法在局部纹理细节和全局整体结构方面的修复均有所提升。

参考文献

- Bertalmio M, Sapiro G, Caselles V, *et al.* Image inpainting. Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques. ACM Press, 2000: 417-424.
- Efros AA, Freeman WT. Image quilting for texture synthesis and transfer. In: Whitton MC, ed. Seminal Graphics Papers: Pushing the Boundaries, Volume 2. New York: ACM, 2023. 59.
- Barnes C, Shechtman E, Finkelstein A, *et al.* PatchMatch: A randomized correspondence algorithm for structural image editing. ACM Transactions on Graphics (TOG), 2009, 28(3): 24.
- Darabi S, Shechtman E, Barnes C, *et al.* Image melding: Combining inconsistent images using patch-based synthesis. ACM Transactions on Graphics (TOG), 2012, 31(4): 82.
- Hays J, Efros AA. Scene completion using millions of photographs. ACM Transactions on Graphics (TOG), 2007, 26(3): 4-es. [doi: 10.1145/1276377.1276382]
- Chen YH. Convolutional neural network for sentence classification [Master's Thesis]. Waterloo: University of Waterloo, 2015.
- Goodfellow I, Pouget-Abadie J, Mirza M, *et al.* Generative adversarial networks. Communications of the ACM, 2020, 63(11): 139-144. [doi: 10.1145/3422622]
- Pathak D, Krähenbühl P, Donahue J, *et al.* Context encoders: Feature learning by inpainting. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 2536-2544.
- Iizuka S, Simo-Serra E, Ishikawa H. Globally and locally consistent image completion. ACM Transactions on Graphics (TOG), 2017, 36(4): 107.
- Hui Z, Li J, Wang XM, *et al.* Image fine-grained inpainting. arXiv:2002.02609, 2020.
- Wang N, Zhang YP, Zhang LF. Dynamic selection network for image inpainting. IEEE Transactions on Image Processing, 2021, 30: 1784-1798. [doi: 10.1109/TIP.2020.3048629]
- Nazeri K, Ng E, Joseph T, *et al.* EdgeConnect: Structure guided image inpainting using edge prediction. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshops. Seoul: IEEE, 2019.

- 3265–3274.
- 13 Ren YR, Yu XM, Zhang RN, *et al.* StructureFlow: Image inpainting via structure-aware appearance flow. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019. 181–190.
- 14 Wu HW, Zhou JT, Li YN. Deep generative model for image inpainting with local binary pattern learning and spatial attention. IEEE Transactions on Multimedia, 2022, 24: 4016–4027. [doi: [10.1109/TMM.2021.3111491](https://doi.org/10.1109/TMM.2021.3111491)]
- 15 Guo ZY, Chen ZB, Yu T, *et al.* Progressive image inpainting with full-resolution residual network. Proceedings of the 27th ACM International Conference on Multimedia. Nice: ACM, 2019. 2496–2504.
- 16 Li JY, Wang N, Zhang LF, *et al.* Recurrent feature reasoning for image inpainting. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 7757–7765.
- 17 Zeng Y, Lin Z, Yang JM, *et al.* High-resolution image inpainting with iterative confidence feedback and guided upsampling. Proceedings of the 16th European Conference on Computer Vision. Glasgow: Springer, 2020. 1–17.
- 18 Yu JH, Lin Z, Yang JM, *et al.* Generative image inpainting with contextual attention. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 5505–5514.
- 19 Dong QL, Cao CJ, Fu YW. Incremental Transformer structure enhanced image inpainting with masking positional encoding. Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022. 11348–11358.
- 20 Quan WZ, Zhang RS, Zhang Y, *et al.* Image inpainting with local and global refinement. IEEE Transactions on Image Processing, 2022, 31: 2405–2420. [doi: [10.1109/TIP.2022.3152624](https://doi.org/10.1109/TIP.2022.3152624)]
- 21 Oktay O, Schlemper J, Le Folgoc L, *et al.* Attention U-Net: Learning where to look for the pancreas. arXiv:1804.03999, 2018.
- 22 Fu J, Liu J, Tian HJ, *et al.* Dual attention network for scene segmentation. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 3141–3149.
- 23 Qin J, Bai HH, Zhao Y. Multi-scale attention network for image inpainting. Computer Vision and Image Understanding, 2021, 204: 103155. [doi: [10.1016/j.cviu.2020.103155](https://doi.org/10.1016/j.cviu.2020.103155)]
- 24 Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 6000–6010.
- 25 Han DC, Pan XR, Han YZ, *et al.* Flatten Transformer: Vision Transformer using focused linear attention. Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision. Paris: IEEE, 2023. 5938–5948.
- 26 Woo S, Park J, Lee JY, *et al.* CBAM: Convolutional block attention module. Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018. 3–19.
- 27 Liu ZW, Luo P, Wang XG, *et al.* Deep learning face attributes in the wild. Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2015. 3730–3738.
- 28 Liu GL, Reda FA, Shih KJ, *et al.* Image inpainting for irregular holes using partial convolutions. Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018. 85–100.
- 29 Guo XF, Yang HY, Huang D. Image inpainting via conditional texture and structure dual generation. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021. 14114–14123.
- 30 Li XG, Guo Q, Lin D, *et al.* MISF: Multi-level interactive siamese filtering for high-fidelity image inpainting. Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022. 1859–1868.
- 31 Deng Y, Hui SQ, Zhou SP, *et al.* T-former: An efficient Transformer for image inpainting. Proceedings of the 30th ACM International Conference on Multimedia. Lisboa: ACM, 2022. 6559–6568.

(校对责编: 张重毅)